SANDIA REPORT

SAND2017-0231
Unlimited Release
Printed September 2016

Cultural Artifact Detection in Long Wave Infrared Imagery

Dylan Z. Anderson, Julia M. Craven, Eric Ramon

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: http://www.ntis.gov/search

SAND2017-0231
Unlimited Release
Printed September 2016

Cultural Artifact Detection in Longwave Infrared Imagery

Dylan Anderson, Julia Craven, and Eric Ramon
Advanced Remote Sensing
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-MS0406

Abstract

Detection of cultural artifacts from airborne remotely sensed data is an important task in the context of on-site inspections. Airborne artifact detection can reduce the size of the search area the ground based inspection team must visit, thereby improving the efficiency of the inspection process.

This report details two algorithms for detection of cultural artifacts in aerial long wave infrared imagery. The first algorithm creates an explicit model for cultural artifacts, and finds data that fits the model. The second algorithm creates a model of the background and finds data that does not fit the model. Both algorithms are applied to orthomosaic imagery generated as part of the MSFE13 data collection campaign under the spectral technology evaluation project.

ACKNOWLEDGMENTS

This work was supported by US Department of State and the U.S. Department of Energy, National Nuclear Security Administration, Office of Nuclear Verification.

CONTENTS

FIGURES

NOMENCLATURE

LWIR	Long Wave Infrared
MSIR	Multi-Spectral Infrared
OSI	On-Site Inspection
CTBT	Comprehensive Nuclear Test Ban Treaty
CTBTO	Comprehensive Nuclear Test Ban Treaty Organization
MSFE13	Multispectral Field Experiment 2013
PSD	Power Spectral Density
RAPS	Radially-Averaged Power Spectrum
SVM	Support Vector Machine
DEM	Digital Elevation Model
MR8	Maximum Response Filterbank 8 [1]

1 INTRODUCTION

This report investigates anomaly detection in aerial, single band, long-wave infrared (LWIR) imagery in the context of on-site inspection (OSI) for the Comprehensive Nuclear Test Ban Treaty (CTBT). The imagery evaluated was collected as part of the Multispectral Field Experiment 2013 (MSFE13)[1] collection campaign conducted by the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). Aerial, remotely sensed imagery (such as the LWIR imagery evaluated in this report) can be collected over large areas with relatively low investment. The ease of collection leads to a data deluge, making it impractical for an analyst to manually review and annotate data in a timely manner. This has created great interest in automated algorithms, which can filter and reduce the amount of data that must be manually evaluated. This report describes two such algorithms for reducing the data burden on analysts.

One of the primary reasons for collecting LWIR imagery is to find evidence of cultural artifacts, which in this report we define as objects in the data that are the result of human activity. Consequently, cultural artifacts include physical structures such as buildings, roads, or vehicles and evidence of activities such as tire marks, holes/mounds of dirt, or disturbed earth. These artifacts are evidence of activity that may require closer inspection to clarify the nature of activities within an inspection area.

The problem considered in this report is the automatic detection of cultural artifacts in LWIR imagery. This can be formulated in two ways. First, by building an explicit model of the artifacts to detect, and finding data points that fit the model. Second, by building a model of the background and finding data points that don't fit the model (otherwise referred to as anomaly detection). This report considers both formulations and proceeds as follows. Section 2 describes the LWIR imagery evaluated. Sections 3 and 4 describe algorithms for automatically detecting cultural artifacts. Section 5 presents results of applying the algorithms to the data. Section 6 summarizes our results and provides recommendations for future work.

[1] The Multispectral Field Experiment 2013 was a PTS airborne MSIR experiment conducted in the Hashemite Kingdom of Jordan.

2 LWIR DATA

The LWIR instrument was part of the MSIR sensor suite included in the MSFE13 data collection. The MSIR system was mounted on a Eurocopter AS332 Super Puma helicopter, and all imagery was collected from a height of approximately 1,400 m above the ground. The LWIR imagery were collected at a resolution of 640x480 pixels. The data is provided in both a binary format (.irb [2]) and as processed png images. The png images are used in this analysis since a reader for the binary format is not available to the authors.

The data appears to be out of focus, as there is substantial blurring in the images. This degradation increases the difficulty of cultural artifact detection. However, data that is collected in focus may appear excessively noisy when processed by the algorithms described in Sections 3 and 4, and may require smoothing be performed prior to analysis.

A more complete and robust background model can be formulated if all of the images are considered jointly. This mitigates local effects in single images such as intensity variation, shapes cut off by image boundaries, etc. To create a better background model and analyze performance in the context of the entire scene, an ortho-rectified mosaic is produced for each of the 5 unique flight lines present in the data. The data as provided does not include any geo-registration information, so a data driven approach for mosaicking the images is taken. This can be done by first computing key points (such as SIFT [3]) on each image, estimating the perspective transform [4] between the images and then stacking the images into a mosaic.

There are a number of commercial software packages available that can be used to perform the mosaicking task. The rise of amateur drones for aerial photography has driven innovation in this space. Commercial packages include Pix4d [5] and Photoscan [6]. There are also open source alternatives (such as Open Drone Map [7]), however these options are not yet as robust or feature complete as the commercial alternatives. The Agisoft Photoscan [6] software package was used to generate the orthomosaics presented in this report. The mosaics for the five unique flight lines are shown in Figure 1. Each flight line is composed of between 400-1,000 individual images. Creating all five orthomosaics requires approximately four hours on an eight core intel i7 workstation.

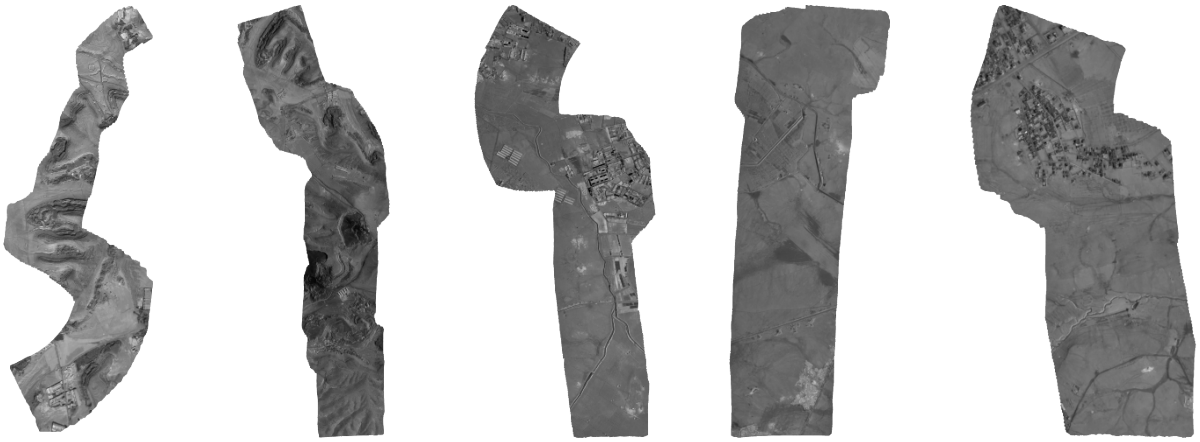


Figure 1. Ortho-mosaic flight lines from the LWIR MSFE13 data collection.

3 TILE BASED FEATURE DETECTION FROM SPATIAL FREQUENCY CONTENT

3.1 Introduction

This section introduces a tile based approach for identifying cultural artifacts in aerial imagery. Imagery is divided into square tiles, and each tile is evaluated and labeled as either suspected of containing cultural artifacts or belonging to the background.

A tile-based approach has the advantage of data reduction: tiles are evaluated and modeled rather than individual pixels. Tiles of size 41 pixels by 41 pixels reduce the number of data points to evaluate by a factor of 1,681. This is of great importance when considering scenes jointly in the form of orthomosaics, which may consist of millions of pixels.

3.2 Motivation

This algorithm starts with the observation that cultural artifacts such as buildings, roads, etc. often contain hard edges and straight lines when viewed in aerial imagery. As such, they contain proportionally more high frequency content. This is illustrated in Figure 2. Figure 2 (a) – (c) contain selections from the first flight line (Figure 1 (a)). Tile (a) illustrates a tile with many cultural artifacts, (b) illustrates a smoothly varying background tile, and (c) illustrates a tile in which this modeling assumption breaks down: a background tile with sharper edges. Figure 2 (d) shows the log power spectrum averaged spatially from each of the three tiles. As expected, (a) features significantly more high frequency content than (b). However, (c) also contains a significant amount of high frequency content, and will likely be confused with cultural artifacts.

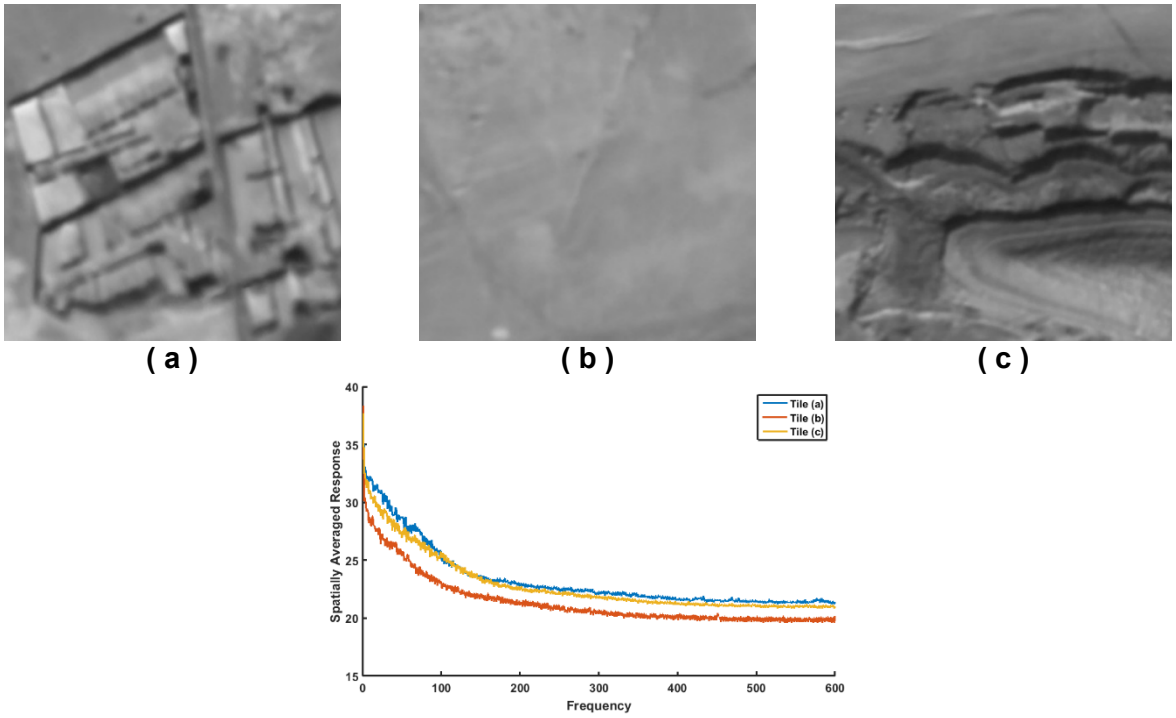


Figure 2. Selected tiles and their radially average power spectra.

3.3 Algorithm Description

The algorithm requires the following user defined parameters:

- τ – tile size. Larger tiles yield reduced computation time and greater spatial frequency resolution at the cost of resolving smaller objects.
- ρ – high frequency cutoff.
- σ – threshold. Increasing leads to reduced missed detections but increased false positives.

Given the modeling assumptions from Section 3.2, the algorithm then proceeds as follows:

1. Tile the input image square tiles of size $\tau \times \tau$.
2. For each tile:
    a. Compute the 2-D power spectral density (PSD) as

    $$S(\nu) = F\{I\}F\{I\}^*$$

    where $F\{I\}$ denotes the Fourier transform of the tile.
    b. Radially average $S(\nu)$ from the center outward to create the radially-averaged 1-D power spectrum, denoted by $RAPS(\omega)$.
    c. Compute the high and low frequency content of $RAPS(\omega)$ by

    $$F_{high} = \sum_{\omega > \rho} RAPS(\omega),\ F_{low} = \sum_{\omega < \rho} RAPS(\omega),$$

    d. Compute the high-frequency proportion as

    $$R = \frac{F_{high}}{F_{high} + F_{low}}$$

3. Create a mask of suspected cultural artifacts by thresholding the high-frequency proportions for each tile as
    $\mathrm{A} = \{tile \,|\, R(tile) > \sigma\}$

3.4 Suggestions for Future Improvements

The final decision made by thresholding the high-frequency proportions is subject to false positives produced by high frequency natural features, such as those shown in Figure 2 (c). A more flexible boundary could be fit by instead binning $RAPS(\omega)$, and treating the resulting bins as a feature vector. This feature vector could then be used in standard classification schemes such as support vector machines (SVM) or logistic regression, provided labeled data were available. This feature vector could also be used in statistical anomaly detection schemes, such as robust covariance [8] [9]. Furthermore, this feature vector could be combined with features from other modalities to reduce false positives. For example, the variability of the local slope from a digital elevation model (DEM) of the tile could be used to reduce false positives introduced by contours such as those shown in Figure 2 (c).

Another way to improve performance is to perform detection in a "multi-scale" sense. Using a permissive (low) value for σ, a scheme to perform a first-pass filtering operation by removing the obviously background tiles would reduce the total number of tiles to consider. Then, a pixel-wise algorithm could be applied on the resulting set. This has two advantages: (1) it reduces the amount of data that must actually be processed on a per-pixel basis, making per-pixel operations tractable and (2) it makes the detection problem more balanced (between cultural artifacts and background), so that standard statistical discrimination algorithms can be more readily applied.

4 TEXTURE BASED ANOMALY DETECTION

4.1 Introduction

This section introduces an alternative model for detecting cultural artifacts. The algorithm presented in Section 3 develops an explicit model of cultural artifacts (that they contain high spatial frequency content). The algorithm presented here takes a different approach: a model is constructed for the *background* and areas not fitting the model are marked as anomalous. This problem framework is often called anomaly or outlier detection, because the outliers exist in the training dataset and it is not known *a priori* which samples are outliers.

The algorithm presented here operates on a per-pixel basis, but could be extended to tiling schemes to reduce computational requirements. Additionally, it can be readily applied to a single image but performs much better in the context of a scene (orthomosaic), as there is more context to develop the background model.

4.2 Motivation

This model starts from a similar observation as Section 3.2: the local texture of the background is significantly different than that of cultural artifacts. This is evident in Figure 2, as tile (a) contains significantly different texture than tile (b) and even tile (c). However, the concept of "texture" is not clearly defined mathematically, and so a proxy representation must be used. The model then makes the following assumption: the background textures are normally distributed, and the anomalous textures have a low probability under the background model distribution.

4.3 Algorithm Description

The algorithm requires the following user defined parameters:

- σ –threshold. Higher values reduce false positives at the cost of missed detections.

Given the modeling assumptions from Section 4.2, the algorithm then proceeds as follows:

1. Compute the local texture vector for each pixel. As a proxy for texture, the Maximum Response (MR) filter bank [1] is applied to the images. This filter bank consists of 38 filters: an edge and a bar filter, each at 6 orientations and 3 scales, and 2 isotropic filters (a Gaussian and a Laplacian of Gaussian). The filters are shown in Figure 3. The final response is computed by taking the max response over the 6 orientations, resulting in 8 responses: an edge and a bar at 3 scales, and 2 isotropic filters. Maximizing over orientation allows for rotational invariance: identical textures have the same response under rotation. The responses to the tile selections from Figure 2 are shown in Figure 4.
2. Flatten the pixel-texture data. After application of the MR8 filter bank, each pixel is represented as a length 8 vector. The resulting data can be stacked into a $pixels \times 8\,responses$ table. To prevent a single response from having an undue

impact on the scene statistics, each column of the table is scaled so that its values lie between 0 and 1:

$$\theta(p,r) = \frac{\theta(p,r) - \min_{p'} \theta(p',r)}{\max_{p'} \theta(p',r) - \min_{p'} \theta(p',r)}$$

where $\theta(p,r)$ denotes the table value for the p^{th} pixel and r^{th} response.

3. Construct a multivariate Gaussian model of the background. Compute the robust covariance matrix using least median regression [8] [9]. This method computes the covariance matrix by minimizing the median of squares instead of the sum of squares as in standard least-squares regression. When developing a model of the background, outliers (anything not part of the true background) should not be used when computing the model parameters as they will dilute the model. Least median regression is robust to up to 50% of the data being polluted as outliers. Least median regression is available in many open source software packages. The implementation provided by the Scikit-learn python package (RobustCovDet, [10]) is used in this analysis.
4. Given the model of the background, measure how well each pixel fits the model via a Mahalanobis distance [11], which can be computed as follows

$$D_M(p) = \sqrt{(\underline{\theta(p)} - \underline{\mu})^T S^{-1} (\underline{\theta(p)} - \underline{\mu})}$$

where $\underline{\mu}$ is the mean response vector and S is the robust covariance matrix.

5. Generate consistent scores that can be compared across scenes by taking the cube root of the distances and scale them to the range $[0, 1]$ to [12]

$$\phi(p) = \frac{\sqrt[3]{D_M(p)} - \min_{p'} \sqrt[3]{D_M(p')}}{\max_{p'} \sqrt[3]{D_M(p')} - \min_{p'} \sqrt[3]{D_M(p')}}$$

6. Threshold the scores to generate a mask of anomalies
   $A = \{p \mid \phi(p) > \sigma\}$
7. Apply binary morphology operations to refine the mask and reduce noise. In this analysis, an opening operation with a disk structuring element followed by a closing operation with a disk structuring element of half the size is performed. Connected components in the resulting mask with area less 200 pixels are removed. This step is highly subjective, and may be considered a post-processing step rather than part of the actual algorithm.

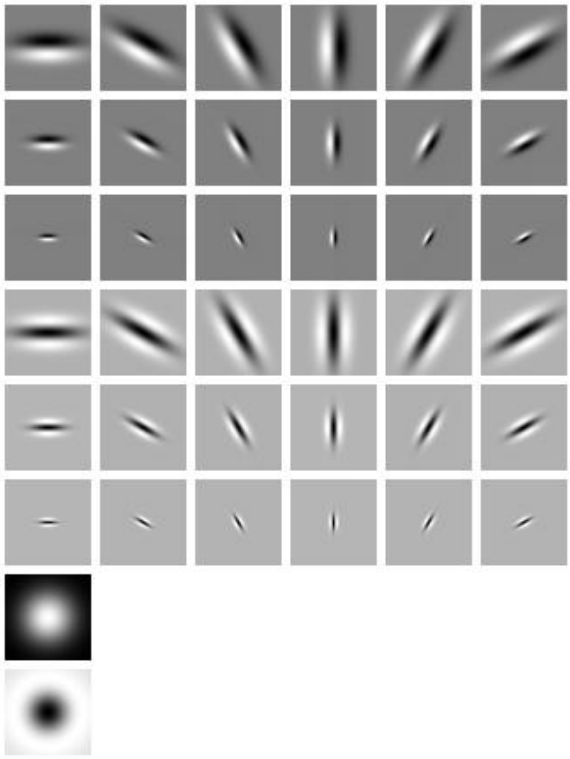
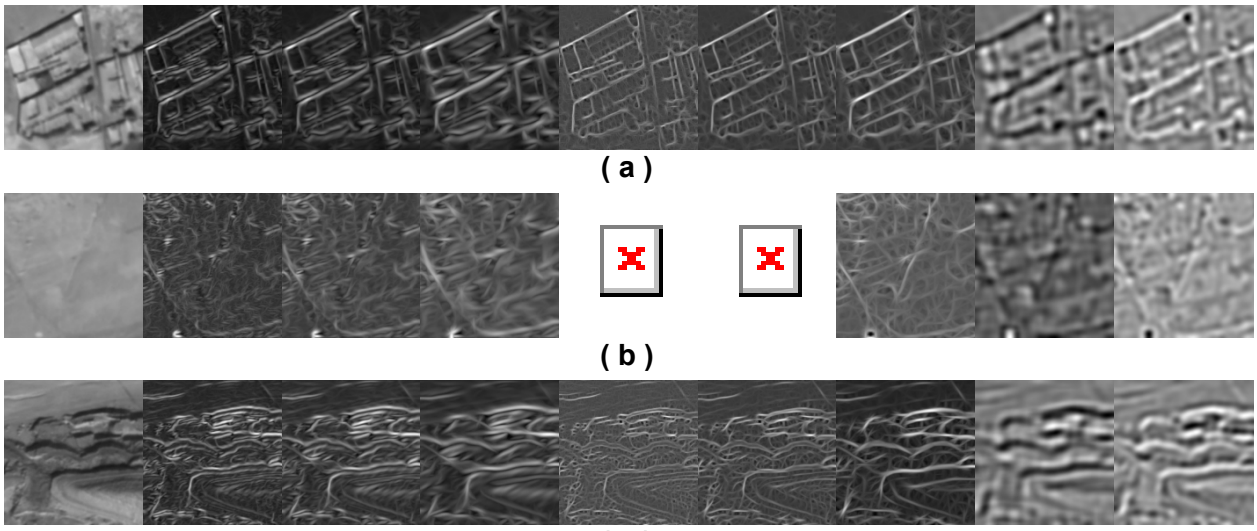


Figure 3. The RFS Filter bank: an edge and a bar filter at 6 orientations and 3 scales, an isotropic Gaussian and Laplacian of Gaussian.



(a)

(b)

(c)

Figure 4. MR8 filter responses for the tile selections from Figure 2.

4.4 Suggestions for Future Improvements

This algorithm introduces two discrete steps that can be combined with other data/algorithms to improve performance: the texture proxy filter responses and the robust Gaussian background model. The filter responses can be combined with other features (such as the binned frequency

vector described in Section 3.4, or other standard image processing techniques [13] [14]), and used with other detection or classification algorithms. The robust Gaussian background model can generically accept any feature vectors, so different input either from other modalities or other feature extraction techniques could be included in the anomaly detection algorithm.

Additionally, the MR8 bank also provides the angle of maximum response for the edge and bar filters across each scale. This information is not used in the analysis described here, but could further enhance the discriminatory ability of the algorithm. The Gaussian background model is simple, and has relatively low computational requirements. As such, it makes a good candidate to operate in a multi-scale type of framework as described in Section 3.4, and would serve as an excellent filtering step prior to discriminating *between* types of anomalies (which is typically a classification problem, instead of a detection as is the focus here).

5 RESULTS

This section presents the performance of the algorithms on the flight line orthomosaics from the MSFE13 data collection. Each flight line is shown separately, with the tile based algorithm from Section 3 shown on the left, and the texture based algorithm from Section 4 shown on the right. It should be noted that detections along the edges are not shown, as the edges are not considered due to the effects of circular convolution when taking the Fourier transform or applying filters (padding cannot be used since the data does not fit neatly in a square).

In all of the flight lines, the texture based anomaly detection algorithm produces much better results than the tile-based model. The computation time of both methods is comparable, with significantly more computation time spent on mosaic creation than either algorithm.

5.1 First flight line

The first flight line contains a large compound near its bottom, a compound on the right edge (that is missed due to edge effects), and what appears to be a small intersection of two roads near the top. It also features several areas of large elevation changes, with exposed rock surfaces. Both algorithms falsely identify these rocky areas as anomalous. The tile-based method has very patchy detections on both the compound and intersection, whereas the texture based method neatly identifies both of these. The overall detections are shown in Figure 5, and close ups of the intersection, compound, and a rocky area are shown in Figure 6.

TILE
Parameters: $\tau = 41, \rho = 0.75, \sigma = 25e-5$

TEXTURE
Parameters: $\sigma = 0.3$

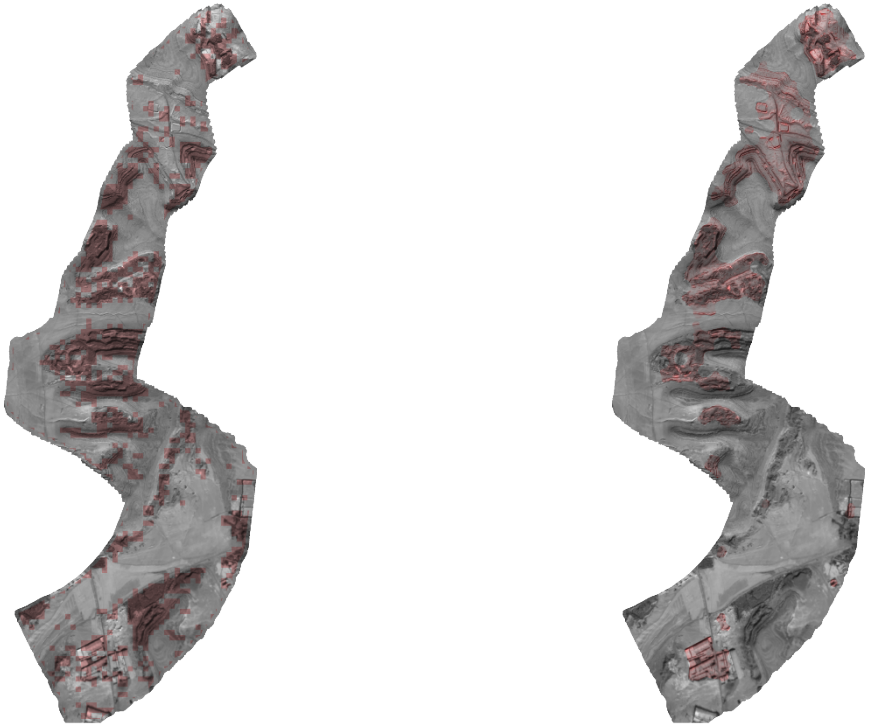


Figure 5. Anomaly detections for the first flight line.

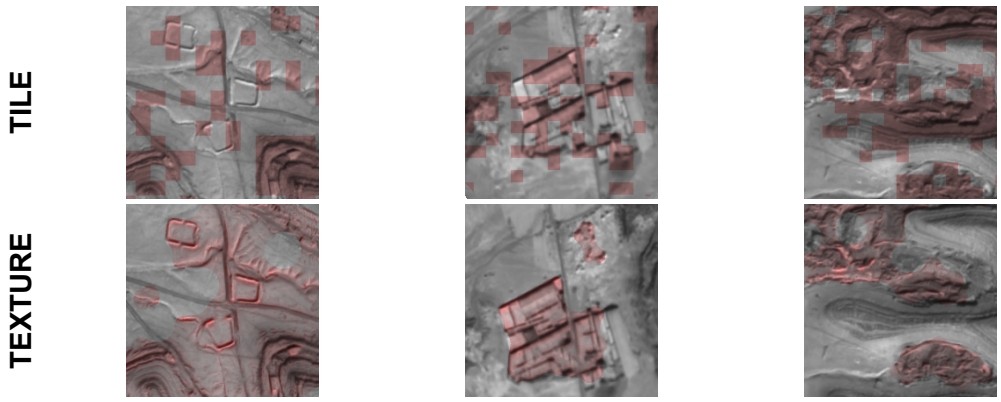


Figure 6. Close ups of the intersection, compound, and rocky area from the first line.

5.2 Second Flight Line

The second flight line features what appears to be several trenches near the top and one near the bottom, as well as exposed rocky surfaces. Again, the rocky surfaces cause false positives for both algorithms, but fewer are present in the texture based method than in the tile based method. The tile based method also has poor detections on the trench areas of interest. The overall detections are shown in Figure 7, and close ups of the trench areas are shown in Figure 8.

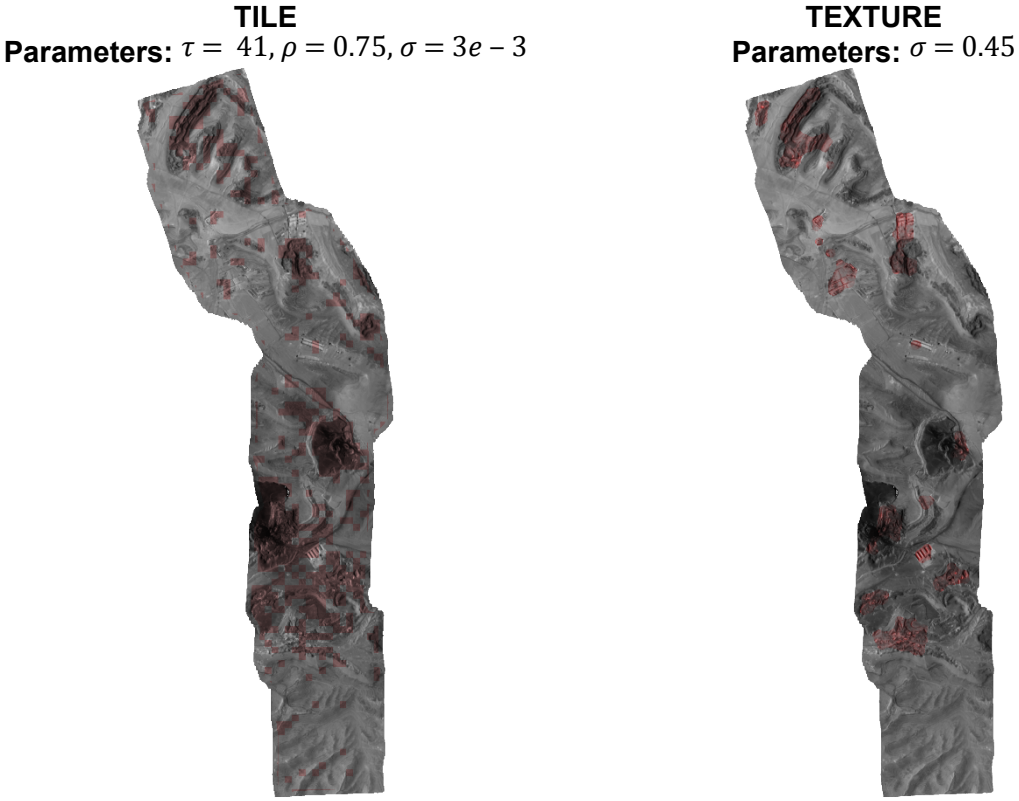


Figure 7. Anomaly detections for the second flight line.

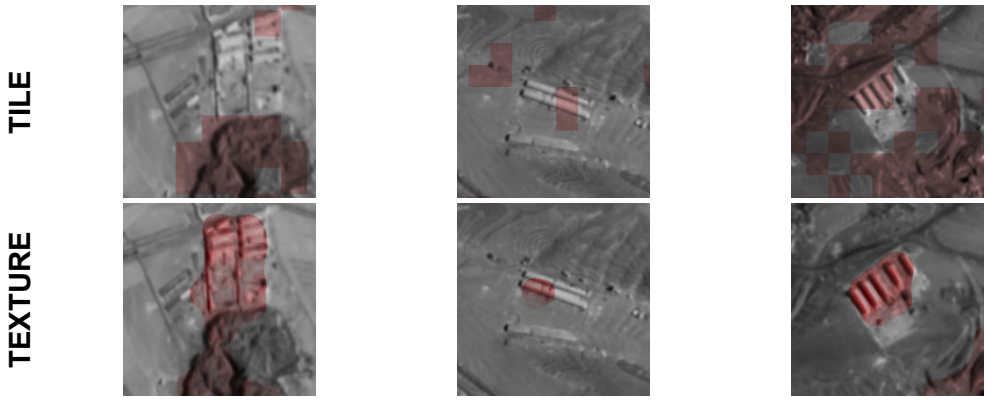


Figure 8. Close up of trench areas from the second flight line.

5.3 Third Flight Line

The third flight line features a larger urban area, several prominent roads, and large open areas. The tiling method still features many false positives, but does a reasonable job of detecting the cultural artifacts. The texture based method neatly fits most of the artifacts, with the primary exception being an inability to detect the thin roads. This is largely the result of the mask cleaning operations applied in post-processing. The overall detections are shown in Figure 9, and close ups of the urban areas and roads are shown in Figure 10.

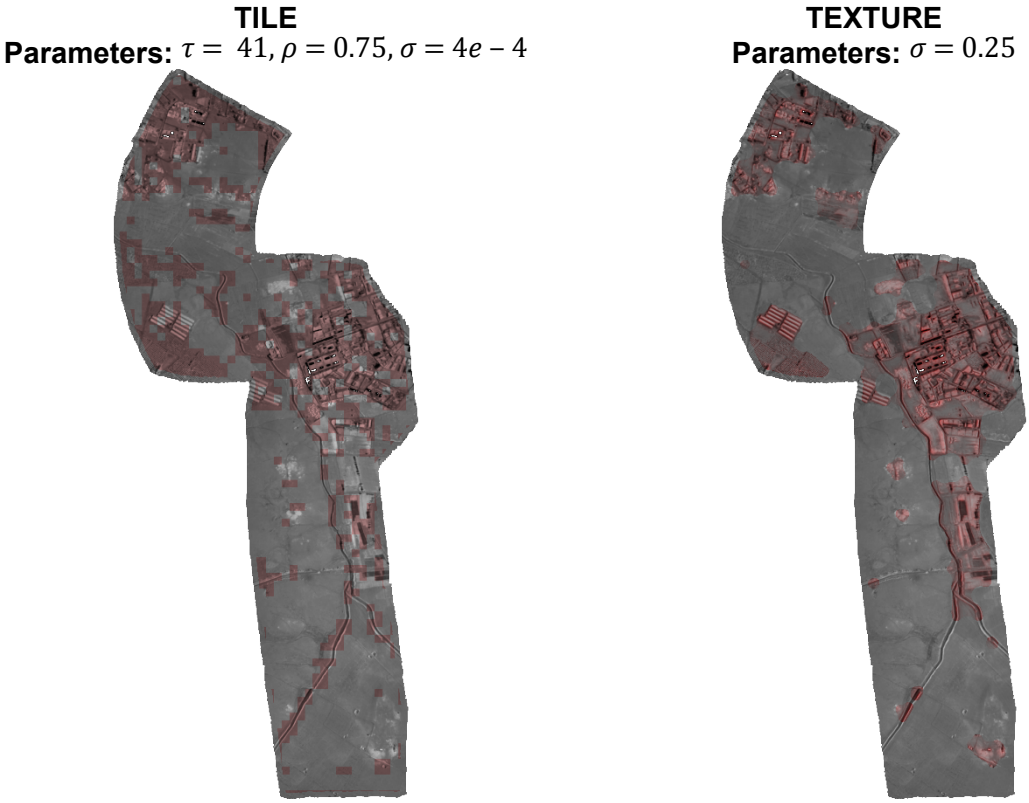


Figure 9. Anomaly detections for the third flight line.

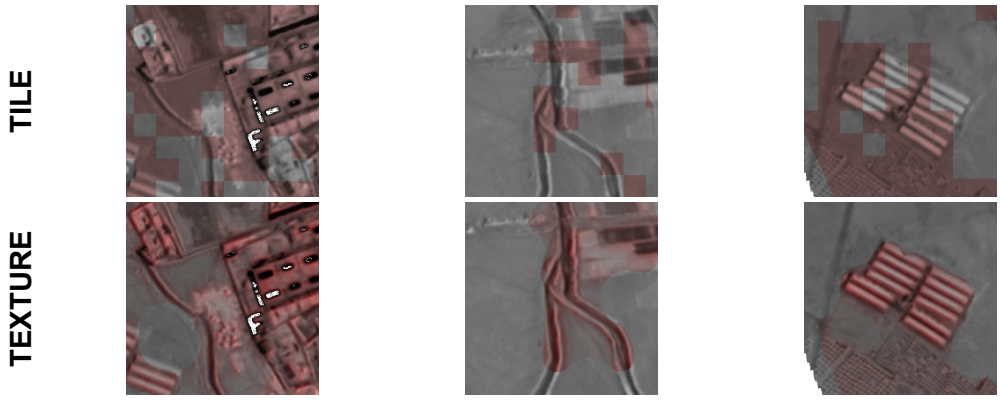


Figure 10. Close up detections of urban areas in the third flight line.

5.4 Fourth Flight Line

The fourth flight line is much sparser than the others, containing a few roads, and an area in which it appears digging has occurred. The texture based method has far fewer false positives than the tiling based method. The overall detections are shown in Figure 11, and close ups of the roads and disturbed earth are shown in Figure 12.

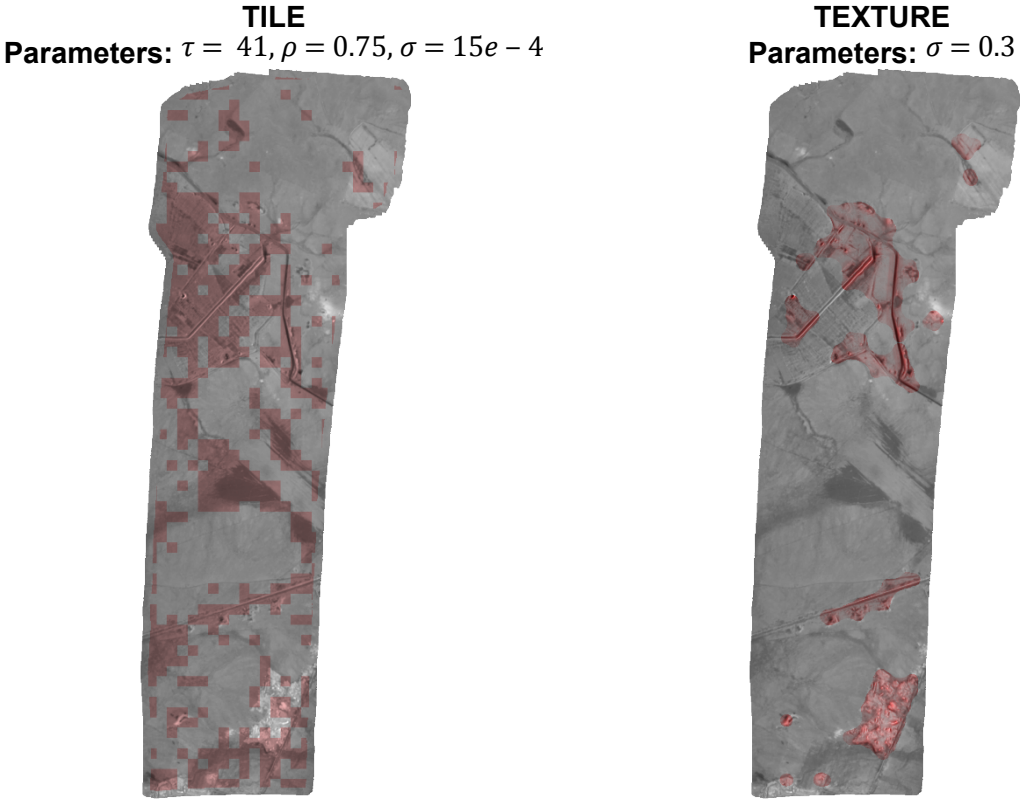


Figure 11. Anomaly detections for the fourth flight line.

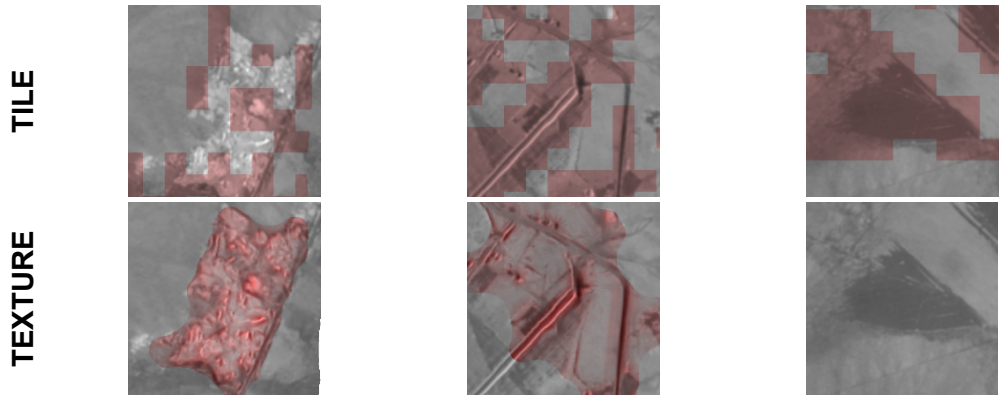


Figure 12. Close ups of disturbed earth, roads, and background for the fourth flight line.

5.5 Fifth Flight Line

The fifth flight line prominently features a large urban area, as well as what appears to be a deep ditch and a large road. As is the case in all of the other flight lines, the texture based method yields much cleaner detections than the tile based method. Both methods do well on the urban area, falsely identify the ditch as an artifact, and poorly identify the road. The overall detections are shown in Figure 13, and close ups of the urban area, ditch, and road are shown in Figure 14.

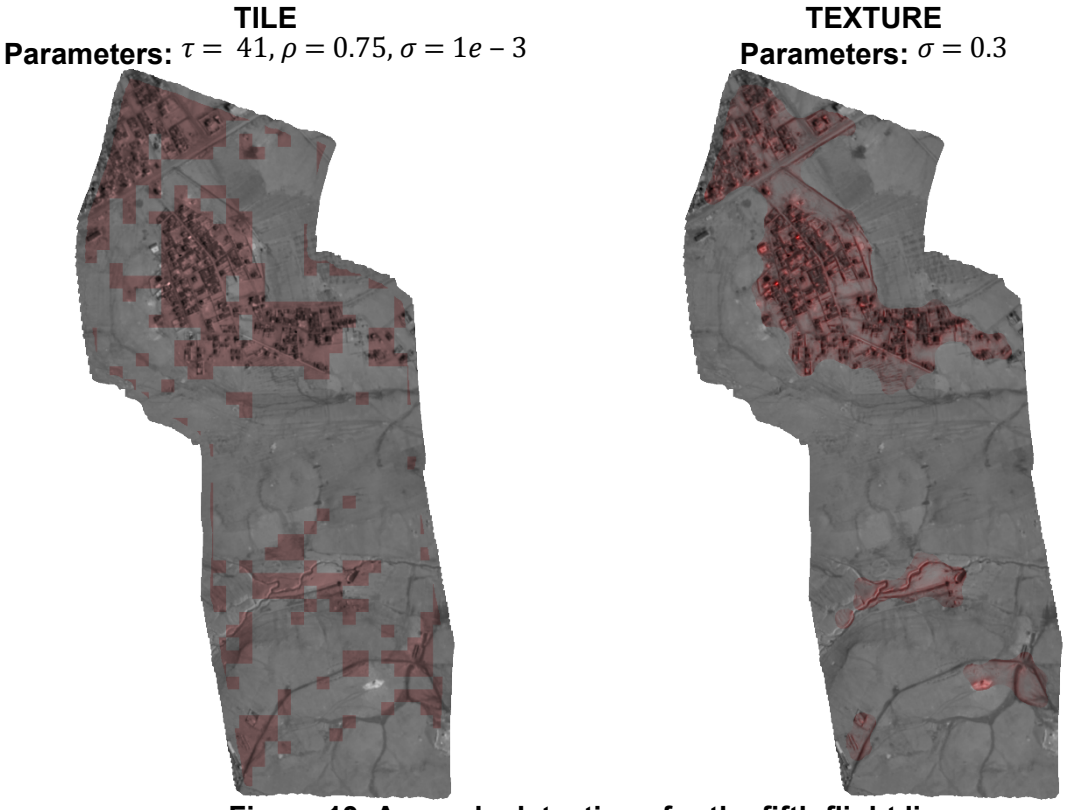


Figure 13. Anomaly detections for the fifth flight line.

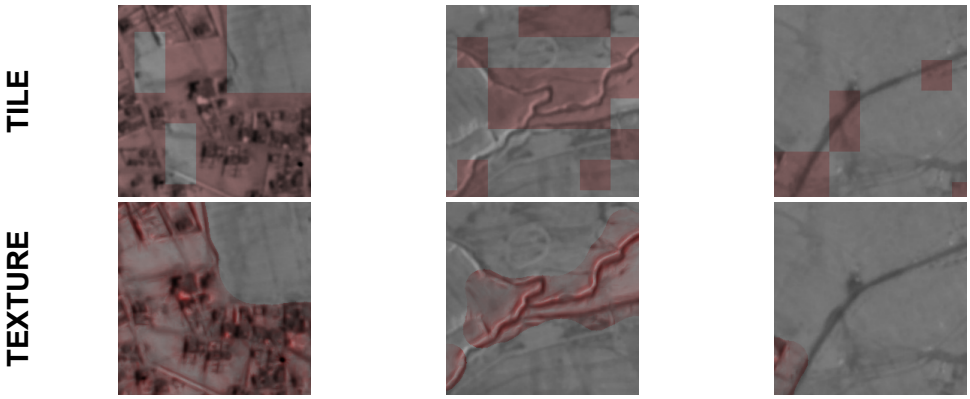


Figure 14. Close ups of the urban area, ditch, and road in the fifth flight line.

6 CONCLUSIONS

This report describes two methods for detecting cultural artifacts in aerial LWIR imagery. The first method operates on sub-tiles of the image and creates an explicit model of cultural artifacts. The second method operates on a per-pixel basis and creates a model of the background, labeling anomalies that don't fit that model. Computational requirements for the two techniques are similar, requiring on the order of 30 seconds to 10 minutes to fit the model on a single core, depending on the size of the input mosaic. It is worth noting that the mosaic step requires significantly more computation, requiring upwards of several hours to compute for a single flight. LWIR data from the MSFE13 collection campaign is stacked into orthomosaics using commercially available software without any geo-location. Orthomosaics allow for data to be considered in the context of the entire scene, and allow for more robust background models.

The texture based anomaly detection scheme presented in Section 4 generally outperformed the tile based method of Section 3. Overall the texture model performed well, but was subject to false positives in rough, craggy areas in which unique rocky texture was present. These false positives could be mitigated by including other data modalities that address these regions. For instance, the variability of the local slope from elevation data could be used to filter many of these false positives.

The techniques presented in this report are very general, and can be readily included as part of a multi-modal joint analysis. It would be valuable to include features extracted from the other collected modalities in the anomaly detection scheme. Additionally, the demonstrated performance of the texture based algorithm suggests that the texture proxy features extracted provide good discriminative information, and would therefore be useful to include in other classification and detection schemes.

The analysis presented in this report should also be performed on the data collected under the IFE14 campaign. This data is very similar in content to the MSFE13 data, but was collected from a much lower altitude. This lower altitude yields images with much less overlap, reducing the ability to generate orthomosaics. The performance decrease resulting from mosaic degradation should be characterized and quantified.

Finally, it would be advantageous to create a rigorous ground truth dataset. Such a dataset could be used to quantitatively evaluate model performance (ROC curves, accuracy, etc). It could also be used to build more powerful supervised models that include the labeling information when constructing a model.

7 REFERENCES

[1] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision,* no. 62.1-2, pp. 61-81, 2005.

[2] Infratec, "Thermographic software," Infratec, 2016. [Online]. Available: http://www.infratec-infrared.com/thermography/thermographic-software.html. [Accessed 31 March 2016].

[3] D. G. Lowe, "Object recognition from local scale-invariant features," *The proceedings of the seventh IEEE international conference on,* vol. 2, pp. 1150-1157, 1999.

[4] I. Carlbom and J. Paciorek, "Planar geometric projections and viewing transformations," *ACM Computing Surveys (CSUR),* no. 10.4, pp. 465-502, 1978.

[5] Pix4D, "Pix4D," [Online]. Available: https://www.pix4d.com/. [Accessed 30 March 2016].

[6] Agisoft, "Agisoft Photoscan," 2016. [Online]. Available: http://www.agisoft.com/. [Accessed 30 March 2016].

[7] "Open Drone Map," 2016. [Online]. Available: https://opendronemap.github.io/odm/. [Accessed 30 March 2016].

[8] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American statistical association,* no. 79.338, pp. 871-880, 1984.

[9] P. J. Rousseeuw and K. Van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics,* no. 41.3, pp. 212-223, 1999.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011.

[11] R. De Maesschalck, D. Jouan-Rimbaud and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems,* no. 50.1, pp. 1-18, 2000.

[12] E. B. Wilson and M. M. Hilferty, "The distribution of chi-square," *Proceedings of the National Academy of Sciences,* no. 17.12, pp. 684-688, 1931.

[13] E. Nowak, F. Jurie and B. Triggs, "Sampling strategies for bag-of-features image classification," *European conference on computer vision,* pp. 490-503, 2006.

[14] C. Yang, C. Wang, Z. Li, L. Zhang and L. Zhang, "Spatial bag-of-features," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 3352-3359, 2010.

DISTRIBUTION

1	MS0899	Technical Library	9536 (electronic copy)